Research Accomplishments and Interests, David Coombs, Ph.D.

If autonomous systems are to realize their potential, they must improve their abilities to operate in the real world. Sensorimotor systems are among the weakest and most critical components of such systems. Perceptual systems exist to enable their owners to cope with the dynamic environments in which they operate. The system is inundated by stimuli; the successful system selects the stimuli that are critically relevant to its current activities and responds appropriately to them. My work is sometimes inspired by principles observed in biological creatures and sensorimotor systems that have stood the test of time. The gaze control work was motivated by the desire to design indexical robot behaviors, which are performed relative to a fixation point or other reference point. The capability for a robot to go-to-the-fixated-object yields robust navigation based on visual feedback despite errors in other senses (e.g., slippage errors in dead-reckoning). My work on low-resolution peripheral vision for mobility was inspired by the honeybee's use of peripheral optical flow to center its flight path through a narrow gap. Reflexive visuomotor behaviors alone cannot achieve our expectations for intelligent robot applications. I am investigating the robot's use of visual attention in building its representations and expectations of its environment. We must verify theories and designs of sensorimotor systems empirically: the world keeps us honest and forces our assumptions to be reasonable. Such experimentation often yields insights and serendipitous discoveries. Integrated systems exhibit interaction effects that cannot be anticipated because they are not easily modeled. The cost of empirical testing must be balanced against the need for rapid prototype development. Nevertheless, the need for testing physical prototypes in challenging environments is clear.

Research Accomplishments

The importance of eye movements to biological visual systems is obvious from their ubiquity. Even insects exhibit eye movements, although they are accomplished by head movements. In contrast, controlled camera movements have played a relatively small role in computer vision research. This disparity presents a rich opportunity for discovery. Biological gaze control systems offer powerful existence proofs of the feasibility of implementing these capabilities and of their utility, but it remains to be learned just how each piece contributes to the organism's behavioral competence. Robots certainly need to employ their visual systems efficiently, and they offer a fertile testing ground for investigating the fundamental role of the perceptual systems in successful behavior.

BINOCULAR VERGENCE ON A TARGET MOVING IN DEPTH

In binocular systems, vergence is the process of adjusting the angle between the eyes (or cameras) so that both eyes are directed at the same world point. Rather than attempting to model primate vergence control in detail, we sought a reasonable implementation of the functional analog for robot heads. Our paper [8] discusses the vergence problem and outlines a general approach to vergence control, consisting of a control loop driven by an algorithm that estimates the vergence error. As a case study, this approach was used to verge the eyes of a binocular robot head in real time.

SMOOTH PURSUIT OF A VISUAL TARGET

My thesis focused on building a smooth pursuit system that holds a moving robot's binocular gaze on a smoothly moving nearby object without requiring the ability to recognize the target in order to distinguish it from distracting surroundings [1]. A novel aspect of our approach is the use of controlled camera movements to simplify the visual processing necessary to keep the cameras locked on the target. A gaze holding system implemented on a binocular robot head demonstrated the approach. Even while the robot moved, the cameras were able to track an object that rotated and moved in three dimensions. The central idea is that localizing attention in 3D space makes precategorical visual processing sufficient to hold gaze. Visual fixation can help separate the target object from distracting surroundings. Converged cameras define a horopter (region of zero stereo disparity) in the scene. Binocular features with no disparity can be located with a simple filter, showing the object's location in the image [3]. Similarly, an object that is being tracked is imaged near the center of the field of view, so spatially-localized processing helps concentrate visual attention on the target. Instead of requiring a way to recognize the target, the system relies on active control of camera movements and binocular fixation segmentation to locate the target.

APPLYING GAZE CONTROL TO ROBOT MOBILITY

At NIST I began applying active vision to mobility. One effort [4] was suggested by the honeybee's use of peripheral motion vision to center itself as it flies through a tunnel or gap [9]. The ability to control egomotion using low resolution peripheral vision is crucial to enable a small high resolution fovea to attend to features that require detailed examination. We have demonstrated the ability of a mobile robot to use low resolution motion vision over large fields of view to steer safely between obstacles. The system uses the maximum flow observed in left and right peripheral visual fields to indicate obstacle proximity. The left and right proximities are compared to steer through the gap. Negative feedback control of steering is able to tolerate inaccuracies in the signal estimation. This simple interpretation of the flows is based on the assumption that the camera is translating along the gaze vector. This condition is maintained under egomotion by active gaze stabilization. Head rotation is countered by opposing eye rotation, and gaze is returned to the heading by rapid nystagmus camera movements when necessary. The low cost of such inexpensive basic navigation competence can free additional resources for attending to the environment.

SAFE MOBILITY USING MOTION VISION ALONE

The lure of using motion vision as a fundamental element in the perception of space drives this effort to use flow features as the sole cues for robot mobility [2]. Real-time estimates of image flow and flow divergence provide the robot's sense of space. Building on the system described above, the robot steers down a conceptual corridor, comparing left and right peripheral flows. Large central flow divergence warns the robot of impending collisions at "dead ends." When this occurs, the robot turns around and resumes wandering. Behavior is generated by directly using flow-based information in the 2-D image sequence; no 3-D reconstruction is attempted. Active mechanical gaze stabilization simplifies the visual interpretation problems by reducing camera rotation. By combining corridor following and dead-end deflection, the robot has wandered around the lab at 30 cm/s for as long as 20 minutes without collision. The ability to support this behavior in real-time with current equipment promises expanded capabilities as computational power increases in the future.

BINOCULAR STEREO TELEPRESENCE OPERATOR INTERFACE

I assembled a prototype wide-angle peripheral and narrow-angle high resolution stereo viewing system for a remote crane operator. We envision moving toward a very natural sense of telepresence for the operator, which will require at least wide screen or wrap-around screen views for peripheral low-resolution vision. I refined the design based on initial experimentation: it involves using multiple camera/lens clusters for each tele-eye to be projected onto a wrap-around screen to provide both high foveal resolution and a wide peripheral field of view to the operator in stereo. Initially the operator will control gaze with a joy-stick or space-mouse. Possible control modes range from simple teleoperation to point-and-click control using a 3D cursor and an indicator of current fixation. Fixation control could be assisted by automatically converging the tele-eyes on the operator’s selected visual target. Our goal is to develop a more natural, transparent, operator interface such as slaving the telehead to the operator’s head movements.

Current Research Interests

INDEXICAL BEHAVIOR AND SERVOING IN VISUAL SPACE

Gaze control and visual feedback can have considerable impact on robust visually-guided behavior in the form of indexical behavior. Developing the capability for a robot to go-to-the-fixated-object could result in robust navigation based on visual feedback even in the presence of other sources of error (e.g., slippage errors in dead-reckoning). The need for precise estimates of the target location and precise calibrations and visuomotor transformations is reduced simply by the use of feedback. An example of robotic behavior that notoriously requires high accuracy in visual perception, transformation and positioning is the task of picking up an object. However, with only coarse calibration, the hand can be carried most of the way toward the target in an initial open-loop reaching phase. A second, visually-servoed, reaching and grasping phase can use visual perception of the hand and the target object in visual coordinates to make relative adjustments of hand position and orientation without requiring transformation to absolute motor coordinates. This approach will be robust with less precise calibration than a completely open-loop system would require.

LOW RESOLUTION VISION

Low-resolution peripheral vision has not been widely applied to low-level vehicle mobility competence. In contrast, it is possible for humans to safely travel a hallway while reading a book or paper (which occupies our scant high-resolution visual sensing) looking up only occasionally to investigate potential threats identified at low-resolution. It must be assumed that much visual support of mobility can be achieved with very low resolution vision. The acuity of human vision drops from 60 cycles/degree at the fovea to 3 cycles/degree at 40 degree eccentricity [7]. This is roughly equivalent to having at least 120 pixels/degree foveally and 6 pixels/degree at 40 degree eccentricity. (To sharpen the comparison, consider that a camera with a 3mm lens that yields a field of view of only 115 degrees will have less than 4.5 pixels/degree in a 512 x 512 image!) Cutting et al. [5] estimate that a human running at 2m=s must estimate her instantaneous heading or within about 4 degrees of its true direction (in order to have sufficient warning to avoid an obstacle). Running at 10m/s requires accuracy of about 1 degree. Interestingly, similar analysis suggests that accuracy of 1 degree is also sufficient for ordinary automobile driving, downhill skiing, and landing an airplane. Duffy [6] reports that primates judging their own heading directions in simulated egomotion displays rely on the visual region between 30 and 60 degrees eccentricity. This strongly suggests that rich visual information for fundamental mobility is found in the visual periphery, and yet acuity is low in the periphery. It appears profitable to consider the use of wider fields of view and lower resolution than are often studied in motion understanding.

FOVEAL-PERIPHERAL VISION AND ATTENTION

In contrast to walking down the hall while reading, it is very difficult (and certainly unsafe) to drive a car with only peripheral vision. The fovea is used to evaluate potential hazards (e.g., peripheral motion that could arise from an object hurtling into the path of the vehicle). Less than 3 degrees of the visual field is covered by the fovea’s high acuity of 60 cycles/degree, and it appears to be sufficient to take foveal samples of the scene 3 or 4 times a second to permit survival in many environments. Clearly it is important to integrate information from these fixations. Perhaps low-resolution peripheral flow navigation and other such primitive capabilities permit us to devote more resources (e.g., time, eye movements, processing power, etc.) to attend to these crucial aspects of the world.

INERTIAL SENSORS

It is essential to understand how to use visual cues, but we must remember that successful creatures combine visual and other cues into a model of target motion and egomotion. Non-visual cues (such as head motions sensed by accelerometers and gyroscopes) can inform gaze stabilization systems as well as contribute to the creature's sense of self-motion. Using such non-visual cues for gaze stabilization and control of self-motion requires models of the system's physical plant so that the proper compensating movements may be made. Furthermore, the appropriate camera movements to compensate for head motion depend on the gaze location, so a representation of the location of objects in three dimensions (or at least depth) is needed. We have equipped a robot with accelerometers and gyroscopes, and we have begun studying their use in actively stabilizing vision in both image processing and mechanical stabilization of the cameras.

INTERACTION WITH THE PHYSICAL WORLD

Systems must become less brittle and be able to operate in less restricted domains. Systems will be more robust and easy to deploy if they can calibrate themselves initially and continually to cope with sensor drift and wear. Further, robots will be more widely useful if they are able to physically interact with their environments in general ways. The ability to open doors and gently push chairs out of the way would open avenues that are closed to a robot that must navigate only in immediately open space. Systems must become more persistent in order to achieve their objectives even in situations that permit success for only some reasonable fraction of attempts.

HUMAN-MACHINE INTERFACES, INCLUDING VIDEO COMMUNICATIONS SYSTEMS

Systems need to be integrated into the everyday human world before we will be able to build fully-competent autonomous systems. We should be deploying systems with limited capabilities to act as assistants as soon as they are able to reduce the burden on humans, and we should work to incrementally increase their functionality as we gain experience. We will benefit not only from improved technology, but also from the experiences of systems interacting with humans in the real world. This has been demonstrated dramatically by the hospital Helpmate of Helpmate Robotics (formerly Transitions Research Corp.), which delivers meals and supplies, thereby freeing the overburdened nursing staff to attend to the medical needs of the patients. The use of these systems requires some accommodation by humans, but it is warranted when the benefit outweighs the cost. The Helpmate is a particularly nice example of this idea, since the hospital environment is modified only modestly, and most of the technology required is fairly general and therefore transferable. Similarly, even slightly intelligent video conference systems could offer a considerable improvement over the crude systems currently in use.

References

[1] David Coombs and Christopher Brown, "Real-Time Binocular Smooth Pursuit," *International Journal of Computer Vision,* 11(2):147-164, 1993.

[2] David Coombs, Martin Herman, Tsai Hong, and Marilyn Nashman, "Real-time obstacle avoidance using central flow divergence and peripheral flow," In *Proc. of ICCV 1995, the Fifth International Conference on Computer Vision*, Cambridge, Massachusetts, June 1995.

[3] David Coombs, Ian Horswill, and Peter vonKaenel, "Disparity Filtering: Proximity Detection and Segmentation," In *Proc. of the SPIE Conf. on Intelligent Robots and Computer Vision XI: Algorithms, Techniques, and Active Vision,* Boston, Massachusetts, November 1992.

[4] David Coombs and Karen Roberts, "Centering Behavior Using Peripheral Vision," In *Proc. of CVPR 1993, the IEEE Conference on Computer Vision and Pattern Recognition,* New York, pages 440-445, June 1993.

[5] James Cutting, Ken Springer, Paul Braren, and Scott Johnson, "Wayfinding on Foot from Information in Retinal, not Optical, Flow," *Journal of Experimental Psychology: General*, 12, 1992.

[6] Charles Duffy, personal communication, 1993.

[7] Susan Galvin and David Williams, "No Aliasing at Edges in Normal Viewing," *Vision Research,* 32(12):2251-2259, 1992.

[8] Thomas Olson and David Coombs, "Real-Time Vergence Control for Binocular Robots," *International Journal of Computer Vision,* 7(1):67-89, November 1991.

[9] M. Srinivasan, M. Lehrer, W. Kirchner, and S. Zhang, "Range Perception Through Apparent Image Speed in Freely Flying Honeybees," *Visual Neuroscience,* 6:519-535, 1991.